

Loading Data

Objective(s)

At the end of this lab you will be able to:

- Read data from a CSV files and load it into database

Exercise

When loading data ifrom a CSV file you need to ensure the data in the dataset maps to the correct datatype and format in the database. One area that can be particularly problematic is DATES, TIMES, and TIMESTAMPs because their formats can vary significantly.

In case the database does not automatically recognize the datatype or format correctly, or the default setting does not match, you will need to manually correct it before loading otherwise you may see an error like the one below when you try to LOAD:

Load details

!

WARNING

0 warning

My computer

Chicago_Crime_Data-v2.csv

Target

QWX76809.CRIME

View Ta

Status

Settings

533

Rows read

0

Rows loaded

533

Rows rejected

Start time

03/15/2020 4:21:43 PM

End time

03/15/2020 4:21:46 PM

The data load job completed with errors.

To investigate, click the **Errors** tab or [load log](#).

Errors 533

0180: The synt
incorrect. SQL
Number of oc

0180: The synt
incorrect. SQL
Number of oc

0180: The synt
incorrect. SQL
Number of oc

0180: The synt
incorrect. SQL
Number of oc

0180: The synt
incorrect. SQL
Number of oc

In order to prevent such errors when loading data, in the Db2 console you can preview the datatype and format of the automatically identified values with the values in the datasets in the LOAD screen such as the one below. If there is an issue, it is usually identified with an Warning icon (red triangle with an exclamation mark) next to the datatype of the column (e.g. DATE column in the example below). To correct this, you may first need to click on the “Clock” icon next to the “Time and Date format” to see the formats, if they are not already visible.

Source
Target
Define

You are loading the file **Chicago_Crime_Data-v2.csv** into **QWX76809.CHICAGO_CRIME_DATA**

Code page (character encoding): **1208 (UTF-8)**

Separator: **,**

Header in first row: ☒

Time format: **MM/DD/YYYY HH:MM:SS**

Date format: **YYYY-MM-DD**

Time format: **HH:MM:SS**

Timestamp format: **YYYY-MM-DD HH:MM:SS**

ID	CASE_NUMBER	DATE	BLOCK
INTEGER	VARCHAR(8)	TIMESTAMP	VARCHAR(35)
1	3512276	HK587712	08/28/2004 05:50:56 PM
2	3406613	HK456306	06/26/2004 12:40:00 PM
3	8002131	HT233595	04/04/2011 05:45:00 AM
4	7903289	HT133522	12/30/2010 04:30:00 PM
5	10402076	HZ138551	02/02/2016 07:30:00 PM
6	7732712	HS540106	09/29/2010 07:59:00 AM
7	10769475	HZ534771	11/30/2016 01:15:00 AM
8	4494340	HL793243	12/16/2005 04:45:00 PM
9	3778925	HL149610	01/28/2005 05:00:00 PM
10	3324217	HK361551	05/13/2004 02:15:00 PM

First check if there is a pre-defined format in the drop down list that matches the format the date/time/timestamp is in the source dataset. If not, type the correct format. Upon doing so, the Mismatch Warning (and exclamation sign) should disappear. In this example below we changed/overwrote the default Timestamp format of ****YYYY-MM-DD HH:MM:SS**** to **MM/DD/YYYY HH:MM:SS TT** to match the value of **08/28/2004 05:50:56 PM** in the dataset.

Code page (character encoding): **1208 (UTF-8)**

Separator: **,**

Header in first row: ☒

Time format: **MM/DD/YYYY HH:MM:SS TT**

Date format: **YYYY-MM-DD**

Time format: **HH:MM:SS**

Timestamp format: **MM/DD/YYYY HH:MM:SS TT**

ID	CASE_NUMBER	DATE	BLOCK
INTEGER	VARCHAR(8)	TIMESTAMP	VARCHAR(35)
1	3512276	HK587712	08/28/2004 05:50:56 PM

Good luck!

Author(s)

[Rav Ahuja](#)

Changelog

Date	Version	Changed by	Change Description
2023-05-04	0.3	Benny	Formatting changes
2020-08-27	0.2	Simran	Migrated Lab to Markdown and added to course repo in GitLab

© IBM Corporation 2023. All rights reserved.